NTU Robotics Final Project: Beverage Servant

Kai Yong Hsieh*, Pin Hao Huang*, Yi Hsuan Lee†, Yu Xiang Tseng†
*Bio-Industrial Mechatronics Engineering, National Taiwan University
†Graduate Institude of Networking and Multimedia, National Taiwan University

***Abstract*—The objective of this project is to build a beverage service robot, we have built voice ordering system, customer detection, and automatic cup picking. Once the Kinect detects the customer raising his/her hand, the position of the customer will be send to the robot and the robot will head to the customer. Amazon Echo Dot is used to receive the ordering commands. And a visual-based controlled arm is used to grab the cup. By combining all the features together, we successfully built an amazing beverage serving bot.**

I. Introduction

Imagine you sitting in your office, feeling overwhelmed under the pressure of work. You're wondering that maybe having a cup of tea or beverage can relieve a bit of your stress, but you just don't want to leave your seat. So, instead, you raise your hand, and immediately a robot come ups to you, takes your order, makes the beverage right away and deliver to your hand. Now, you have a tasty beverage without spending any efforts. It sounds quite awesome, isn't it? Believe or not, we are making it true.

In our final project, we built a robot beverage servant which can serve four kinds of beverage, water, milk, black tea, and milk tea. We have a Kinect that was seperated from the robot and makes a global view of the environment. As soon as it detects the customer who is raising his/her hands, it demands the robot to approach that customer. Next, the customer will make the ordering by talking to the Amazon Echo Dot. After the customer finish the ordering, the robot arm will grasp the cup, filled with the beverage by the pumps and tubes, and finally deliver it to the customer.

We describe our overall architecture in Section III-A. The whole architecture contains five main parts:

- To locate where the customer is, we install a Kinect on a tripod. We made it calculate the displacement from the Kinect itself to the customer, and also the angle from the straight view of the Kinect to the customer. For more details, see Section III-E.
- We build our voice ordering system on Amazon Echo Dot. The customer can easily start the ordering service by calling: *Alexa, drink service!*. The details are showed in Section III-C.
- The robot arm plays an important role in serving beverages to customers, since it has to pick up the cups, and hand it to the customers. In order to achieve this, we build our robot arm controller from scratch based on Arduino. The details of the forward and inverse kinematics are described in Section III-B.
- The customer will randomly place the cup anywhere on the platform. So we install a camera on the robot arm to make it able to locate where the cup is. The procedure of this part is described in Section III-D.
- Finally, after the arm picks up the cup, the pump starts automatically to inject the beverage into the cup. This process is described in detail in Section III-F.

II. Related Work

We found some amazing works that are similar to our project. One is the BoozeBot[1], a cocktail-making robot. BoozeBot uses a touch screen to let user pick the drink they want. Then, the robot uses multiple pumps to mix the drink. Our robot uses the similar approach for mixing the drink. However, the differences are we can let the robot move to us and ask the drink by voice commands. Another similar work is Techi Casino[2], a drinks delivery robot. This robot moves around casino and stops when the button is pressed. However, the robot doesnt know which drink the user picked and also doesnt know when to stop by the users location. Our robot not only moves to the user location precisely but also gives the desired drink to the user.

Fig. 1: BoozeBot

Fig. 2: Techi Casino

III. Methodology

A. Architecture

B. Robot arm control

The robot arm we use in this project is Uarm, which the schematic diagram of the arm is showed in Figure 1. The arm is controlled by three servo motors and are all installed in the origin, one controls the angle of A directly, the other one controls the angle of B indirectly by linkage C, and the third one controls the rotation of the entire structure about Z axis. This ingenious structure enables the end effector to face downward vertically in any configuration. The whole structure can be simplified as two connected links showed in figure 4.

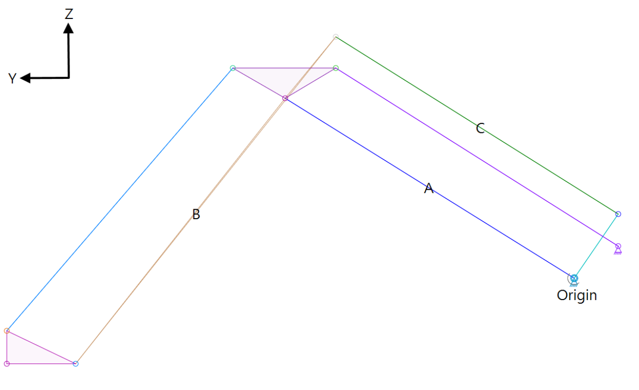


Fig. 3: Robot arm.

For the forward kinematics of the arm, we define θ_A, the angle between link A and the Y axis. θ_B, the angle between link B and the Y axis. θ_C, the angle between link A and the X axis. ℓ_A, the length of link A. ℓ_B, the length of link B. $\mathcal{D}$, the horizontal distance from the origin to the end effector which is given by:

$$\mathcal{D} = \ell_A \cos(\theta_A) + \ell_B \cos(\theta_B) \tag{1}$$

The x and y coordinates can be given by the transformation from the Polar coordinate system to the Cartesian coordinate system:

$$\begin{aligned} x &= \mathcal{D}\cos(\theta_C) \\ y &= \mathcal{D}\sin(\theta_C) \end{aligned} \tag{2}$$

And the z coordinate can be given by:

$$z = \ell_A \sin(\theta_A) - \ell_B \sin(\theta_B) \tag{3}$$

The inverse kinematics are simple. First, θ_C can be solved by:

$$\theta_C = \arctan(\frac{y}{x}) \tag{4}$$

θ_A and θ_B can be solved by simple geometry, which is showed in Figure 5:

$$\theta_A = \arccos(\frac{\ell_A^2 + (z^2 + \mathcal{D}^2) - \ell_B^2}{2\ell_A\sqrt{z^2 + \mathcal{D}^2}}) + \arctan(\frac{z}{\mathcal{D}}) \tag{5}$$

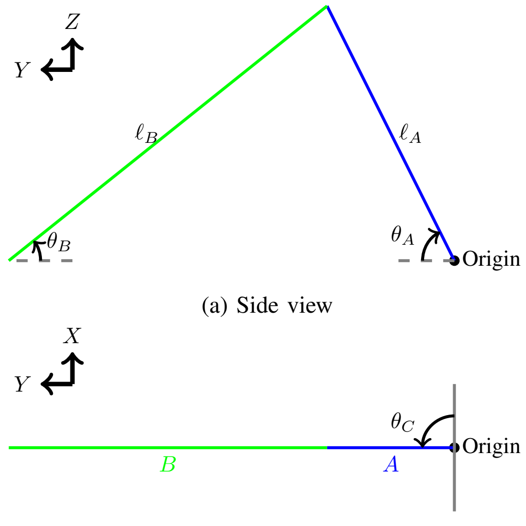
$$\theta_B = \arccos(\frac{\ell_B^2 + (z^2 + \mathcal{D}^2) - \ell_A^2}{2\ell_B\sqrt{z^2 + \mathcal{D}^2}}) - \arctan(\frac{z}{\mathcal{D}}) \tag{6}$$

where $\mathcal{D}$ can be given from Equation (2):

$$\mathcal{D} = \sqrt{x^2 + y^2} \tag{7}$$

So, with ℓ_A and ℓ_B known, given x, y, z, we can easily get the desire configuration: θ_A, θ_B, θ_C.

The code is written in C++ and the arm is controlled by Arduino.



(a) Side view

(b) Top view

Fig. 4: Simplified illustration

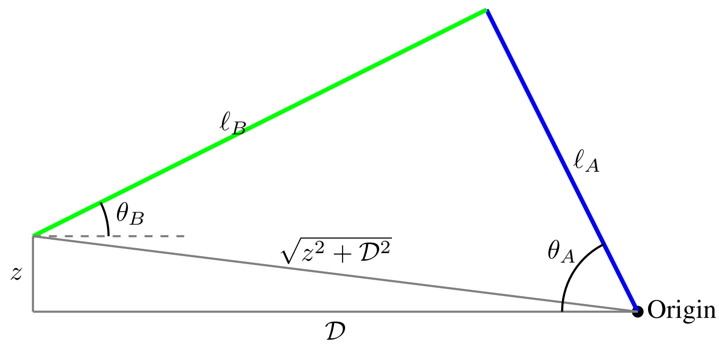


Fig. 5: The geometry relationship

C. *Voice ordering system*

We build our voice recognition system on the Amazon Echo Dot. The dialog example is showed in Figure 6. To develop the system, first, we configured a custom Alexa Skill called *Drink Service* by using the Alexa Skill Kit[3], here we define the requests the skill can handle (*intents*) in a JSON structure and also the words (utterance) users may say to invoke those requests. We then build our code based on the AWS (Amazon Web Services) Lambda function, which is a cloud-based service that handles the requests for the skill type and host it in the cloud. The code was written in Node.js, it handles the request and then figure out the corresponding response and send back to Echo Dot. If the request isn't valid, the system will send a response and ask the user to order again.

For example, if the user order something that doesn't exist in the menu, the system will inform the user that his/her choice isn't valid and will ask the user to choose from the list of our beverage. As long as the system recognize a valid order, it will send a "successful ordering message" back to Echo Dot, and also sends a HTTP request that contains the data to the Nvidia TX2. We The flowchart of our system is showed in Figure 7.

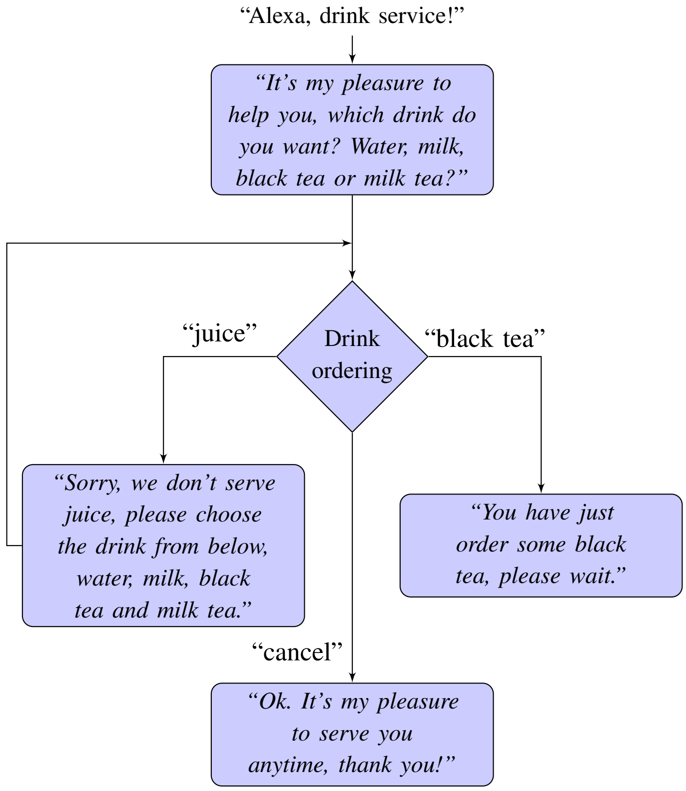


Fig. 6: Dialog example

D. Cup detection and grab

Considered the customer may place the cup anywhere on the platform, automatically locate and grab the cup is necessary. We use visual feedback controlling to address this task. A camera with 130 degree wide angle lens was installed in front of the claw to get the whole view of platform. To make the cup detection easier, we pasted blue tapes around the edges of the cup. In image processing, we first convert the color space of the image from BGR to HSV, and extract only the blue region in the image. We then use the Hough circle algorithm to find the center of the cup. We define the target position $\mathcal{P}$, which is the position of the center of the cup in the image when the claw is right above the cup.

Since the camera is installed in front of the claw, there's a bias distance between the center of the claw and the center of the camera. We address this issue by adding a constant to the target position to be our final target position $\mathcal{P}^*$. After the displacement Δd between the final target position and the

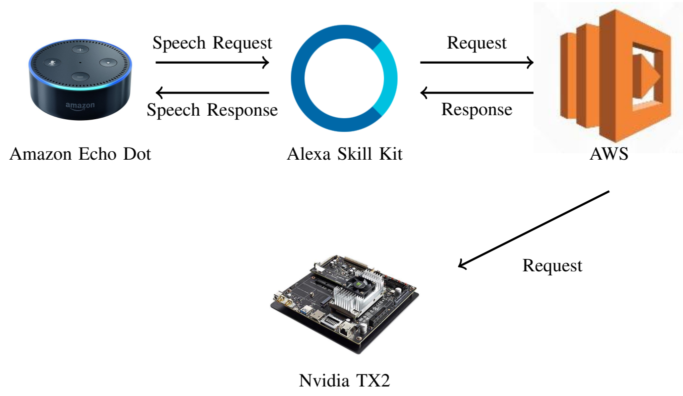


Fig. 7: Alexa Skill setup and the integration with Nvidia TX2

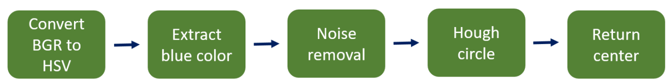


Fig. 8: Image Processing

center of the cup is calculated, we convert it into the real world distance through simple linear transformation:

$$\Delta d_R = C_1 \Delta d + C_2 \tag{8}$$

Which C_1 and C_2 are some constants. The robot arm will move toward the direction that gradually decrease Δd_R until the center of cup meets the final target coordinate.

E. Hands-up person detection

We use Kinect to detect the customer that raise their hand to call for service. We used Body Tracking which is one of the features in Kinect. This Windows library[4][5] provides access to the the information of any individual bodies that Kinect detects. For any individual body, we could get the information such as the position and the orientation of the body joints. The position refers to the 3D coordinate system defined in (x, y,

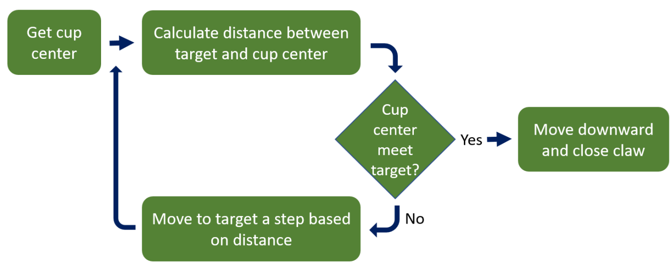


Fig. 9: Cup grasping flowchart

z), which the origin (x=0, y=0, z=0) is located at the center of the IR sensor on Kinect. The X axis points toward the left side of the sensor, the Y axis points upward, and the Z axis points to the direction which the sensor is facing. The unit is in meters. In this work, we tracked the three body joints position: *head*, *right hand*, and *left hand*. When the Kinect detect that the y coordinate of the *left hand* or *right hand* is greater than the one of the *head*, we indicate this as a *call for service* signal and then sent the distance, $\sqrt{x^2+z^2}$ as well as the angle, $\arctan(\frac{x}{z})$ of the *head* to the Nvidia TX2 to start the service.

F. Beverage pumping

For beverages, we allowed user to pick 4 kinds of drinks: water, milk, black tea, and milk tea. Since milk tea is combined by black tea and milk, we only need 3 pumps for the robot. We placed 3 nozzles at the center of the robot arm in order to mix the drink in the cup. Then, we use soft tubes to connect them. We tried out few experiments to find out how to control the pump to get the desire amount of drink. Based on multiple trials, we figured out that we need to pump the drink for 6 seconds to get half of the cut full. But how about a full cup? By common sense, we need to pump for 12 seconds to get a full cup. However, in practice, pumping for 12 seconds would be more than the desired amount. Since, during the 6 seconds for pumping half a cup, the motor inside the pump goes through three states: acceleration, fully-rotated pumping, and deceleration. The volume of acceleration and deceleration is constant. But the fully-rotated pumping time varies each time. In order to fill a cup, we add a 1 second stop in the middle of 2 half-cup iterations. In this way, we easily solved the pumping issue.

IV. Experimental Set up

- Origin - We define out origin as the location of Kinect,which is also the starting point of our robot, so that the angle and distance of customer Kinect detected can be directly convert and send to robot.
- Arm - The initial position of the arm is about 30 cm higher above the platform, so that the camera can get the whole view of platform.
- Kinect - Since our robot has no place to install Kinect and it requires the power supply, it was set up on a tripod and connected to the power supply cables and is connected to a laptop.
- Pumping System - Since we placed our arm in the front side of the robot, we set the pumping system in the back of the robot in order to keep it balanced.

Our beverage serving robot is showed in Figure 10. The whole experiment set up is showed in Figure 11.

V. Results

Integrating all the works we've done, we successfully made this drink servant serve drinks to customers during the in-class demonstration. However, there are many parts that we can make it better.

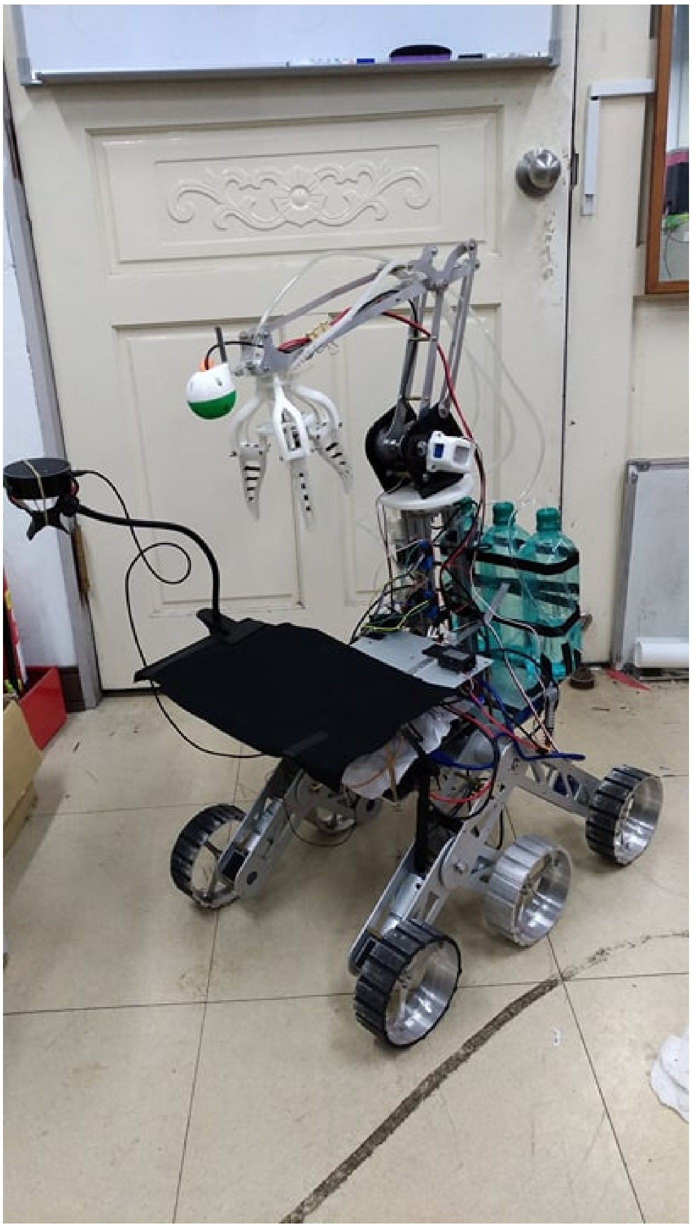


Fig. 10: Our beverage serving bot

- There are some limitations of the Kinect, which we can't detect the customers that are farther than 4.5 m, or sometimes there are some delays of the detection. Initially, we'd planned to install the Kinect sensor on the robot, so that it can track the customer by real-time feedback, and can also be able to avoid obstacles. However, we have limited time and their is no space left for us to install it on the robot, so we instead set the Kinect on a tripod which eventually simplified our task but also creates more limitation for the ability of our robot. And since we detect the customer only in one shot, the robot couldn't always precisely arrived to the position of the customer.
- The brightness of the environment affects a lot to the

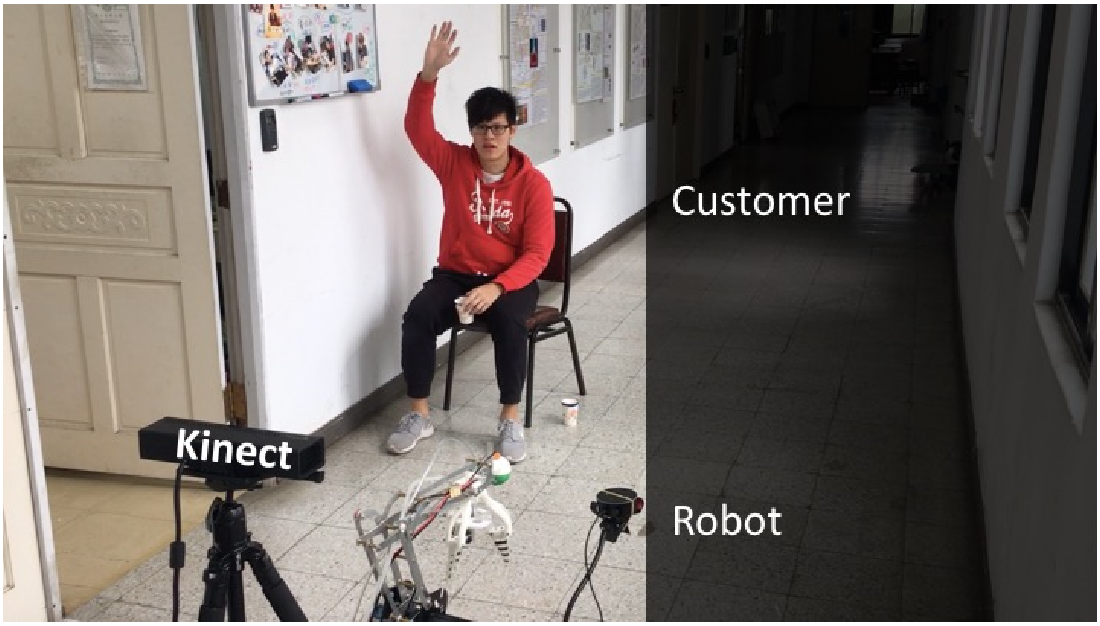


Fig. 11: Experiment set up

success of picking up the cup. To dark or to bright will lead to failures. Moreover, the robot arm isn't precise enough and will somehow affect the accuracy to grab the cup.

- We didn't consider the varying viscosity and density of different beverage, which the pump only fills the tube while fails to fill the cup since it isn't powerful enough.

VI. Conclusion

In this final project, we integrated lots of features and finally came out with this amazing beverage serving bot. We built customer detection, voice ordering, and automatic cup catching. As a matter of the fact that there are still many things for us to improve, nonetheless, we undoubtedly did the prototype of the beverage service robot that apply the knowledge we learned in the class to this work and we would try to make it more useful to build a powerful and complete beverage service robot in the future.

VII. Work division

The team members and ID are showed in Table I. The work division of this project are showed in Table II.

TABLE I: Team members

Members	ID
Kai-Yong Hsieh	B03611041
Pin-Hao Huang	B03611032
Yi-Hsuan Lee	R06944014
Yu-Hsiang Tseng	R06944024

TABLE II: Work Division

Task	Contributor
Robot mechanical restructuring	Kai-Yong Hsieh / Yu-Hsiang Tseng
Robot arm controller	Pin-Hao Huang
Voice ordering system	Pin-Hao Huang
Cup detection	Kai-Yong Hsieh
Robot arm manipulation	Kai-Yong Hsieh
Hands-up person detection	Yi-Hsuan Lee
Beverage pumping	Yu-Hsiang Tseng / Pin-Hao Huang
Echo Dot and Nvidia TX2 communication	Pin-Hao Huang / Kai-Yong Hsieh
Kinect and Nvidia TX2 communcation	Yi-Hsuan Lee
Overall integration	Kai-Yong Hsieh

VIII. Reference

[1]https://laughingsquid.com/boozebot-a-cocktail-making-robot-that-can-mix-over-4500-different-dri

[2]http://techmetics-group.com/techi-casino/

[3]https://developer.amazon.com/docs/ask-overviews/build-skills-with-the-alexa-skills-kit.html

[4]https://msdn.microsoft.com/en-us/library/dn799273.aspx

[5]https://kheresy.wordpress.com/2015/03/03/k4w-v2-part-7-user-skeleton/

IX. Video Link

https://www.youtube.com/watch?v=tUzVZfe1XYk&feature=youtu.be